



Curso: ¿Cómo se maneja SPSS como un profesional? Introducción al trabajo con sintaxis

José Manuel Torrado
Profesor Ayudante Doctor
Departamento de Sociología, Universidad de Granada
josetr@ugr.es

LECTURA 1: EL PROCESO DE INVESTIGACIÓN SOCIAL CUANTITATIVA

En las páginas que siguen trato de sintetizar el proceso de investigación social cuantitativa paso a paso. La finalidad es que el alumno adquiera una visión global del proceso y, especialmente, del papel que tiene en el mismo el análisis de datos, de manera que no entienda el curso de manera aislada y pueda ver las herramientas que vamos a trabajar (SPSS) dentro del proceso más amplio en el que se enmarcan, ya que es la única forma de comprender su utilidad real. Además, a lo largo del texto se dan algunas claves básicas sobre el proceso de análisis de datos, las cuales son fundamentales para seguir el curso de manera satisfactoria. A continuación pasamos a ver el complejo proceso de investigación social cuantitativa paso a paso.

1) Formulación de un objetivo de investigación.

Es el primer, e ineludible, paso para poder iniciar cualquier tipo de investigación social. Consiste en que nos planteemos un objetivo, algo que queremos conocer, explicar, describir, analizar… La formulación de un objetivo de investigación debe seguir algunas reglas básicas. Primero, el objetivo debe ser una proposición lingüística, una frase corta y clara, que prácticamente cualquiera pueda comprender. Segundo, dicha frase siempre debe comenzar por un verbo en infinitivo, los más usuales en investigación cuantitativa son: conocer, describir o analizar. Tercero, el objetivo no puede incluir juicios de hecho o valor, es decir, debe excluir aseveraciones a priori sobre la realidad que vamos a estudiar (juicios de hecho) y valoraciones morales sobre la misma (juicios de valor). Por ejemplo, un objetivo que incluye un juicio de hecho sería: "Conocer por qué los jóvenes españoles se emancipan tan tarde", el afirmar que se emancipan tarde es un juicio de hecho sobre algo que aún desconocemos, sería necesario primero determinar la edad de emancipación y, por comparación con otros países del entorno, determinar si efectivamente se emancipan tarde. Una forma correcta de enunciar el objetivo sería: "Conocer la edad de emancipación de los jóvenes españoles" o "Conocer las diferencias entre la edad de emancipación de los jóvenes de España respecto a otros países del entorno". Por último, el objetivo debe ser contrastable con los medios disponibles. Conviene alejarse de grandes preguntas como ¿Qué explica la desigualdad social en las sociedades capitalistas?, al menos a priori, son preguntas difícilmente contrastables y, aunque pueden abarcarse (Marx intentó aproximarse a explicar el funcionamiento del capitalismo) son demasiado ambiciosas para un investigador novel, además su resolución implica hacer preguntas más pequeñas que permitan ir construyendo ese objetivo más vasto. En cuarto lugar, es conveniente concretar el objetivo en un espacio-tiempo determinado, es decir, acotarlo a aquella parcela de la realidad social a la que

podemos tener acceso. Por último, cabe destacar que un objetivo de investigación no es otra cosa que una pregunta que el investigador le hace a la realidad social y que pretende responder recogiendo y analizando datos, por ello, un buen truco para saber si un objetivo de investigación está bien formulado es transformarlo en pregunta, si puedes transformar el objetivo formulado en una pregunta sin apenas esfuerzo, esto es una señal de que el objetivo está bien formulado.

Tabla 1. Ejemplos de objetivos de investigación
• Conocer qué factores explican el rendimiento académico del alumnado de Sociología de la UGR • Analizar los determinantes de los movimientos hacia la corona metropolitana de Granada • Explicar la decisión de votar a partidos de extrema derecha • Analizar el impacto que tuvo el confinamiento en la salud percibida de la población

2) Formulación de hipótesis.

Una vez tenemos el objetivo de investigación formulado debemos formular las hipótesis de investigación. Las hipótesis no son otra cosa que respuestas temporales al objetivo de investigación enunciado, a la pregunta planteada. Por ejemplo, si mi objetivo de investigación es “Conocer qué explica el rendimiento académico de los estudiantes del primer curso del grado en Sociología de la UGR”, una hipótesis podría ser “El hecho de ser mujer implica un mayor rendimiento académico”. Como veis, una hipótesis no es otra cosa que una frase corta, clara y concisa que pone en relación el problema o aspecto de la realidad que quiero conocer (la variable dependiente, en este caso el rendimiento académico), con sus factores explicativos (la o las variables independientes, en este caso el sexo de la persona – ser mujer –).

Pero ¿cómo construyo una hipótesis? ¿En qué debo basarme? No basta con hacer una aseveración, así sin más. Si bien algunos expertos (muy consolidados en un campo de estudio) pueden permitirse el lujo de formular hipótesis según su criterio; el común de los mortales revisa la literatura científica previa sobre la cuestión, repasando de manera sistemática todos los estudios que se han planteado nuestro mismo objetivo de investigación (o al menos, uno similar), y construye las hipótesis sintetizando las conclusiones a las que llegaron.

Unas preguntas muy pertinentes que los estudiantes suelen plantear cuando hablamos de hipótesis es ¿realmente son necesarias? ¿Por qué debo dar una respuesta a mi pregunta de investigación antes de extraer y analizar datos? La respuesta es sencilla, y es que las hipótesis son el mecanismo por el cual la ciencia avanza en el conocimiento de la realidad. La ciencia es un saber acumulativo, fundamentado en conocer la realidad a través del análisis empírico – recoger y analizar datos – para así construir explicaciones

(teorías) sobre cómo funciona la realidad, siempre y únicamente basadas en los datos disponibles. Las explicaciones o teorías construidas por los científicos que han estudiado nuestro objeto de investigación antes que nosotros deben ser contrastadas para comprobar si siguen estando vigentes (se confirma la hipótesis), o por el contrario el investigador se equivocaba, o simplemente la realidad ha dejado de funcionar de la manera que él constató (se rechaza la hipótesis). Poner en cuestión aquellos hechos que fueron constatados por otros científicos y someterlos a comprobación empírica es la forma que tiene la ciencia de avanzar hacia cotas cada vez mayores de conocimiento de la realidad. Negarse a sí misma, cuestionar las explicaciones construidas para poder construir cada vez mejores explicaciones (teorías) del mundo, es el objetivo de la ciencia. Esto es lo que el filósofo austriaco Karl Popper llamaba el principio de falsabilidad, un principio básico sin el cual la ciencia quedaría reducida a mera doxa o peor, a un dogma.

3) Operacionalización de conceptos y determinación de las variables clave.

Una vez formulados objetivos e hipótesis debemos operacionalizar los conceptos que hemos utilizado en su enunciación. ¿Opera… qué? Soy consciente que la palabra operacionalización suena rara a aquellos que no se dedican a la investigación social, sin embargo es un concepto relativamente simple. Macionis y Plummer (2007), en su manual de introducción a la Sociología, nos dicen que "operacionalizar es definir exactamente lo que se quiere medir". Es decir, operacionalizar no es otra cosa que transformar conceptos abstractos en variables susceptibles de medición. Dicho de otro modo, transformar nuestros conceptos de manera que podamos darle una transcripción o codificación numérica para, posteriormente poder medirlos cuando vayamos al campo. Una variable operacionalizada puede fácilmente medirse y, cuando nuestra unidad de análisis sean individuos (lo más común en ciencias sociales), podemos transformarla en una pregunta que el individuo puede contestar, de manera que podemos medir tal o cual característica, sentimiento u opinión del individuo. En la tabla inferior os propongo algunos ejemplos de conceptos relativamente complejos y propuestas de operacionalización, creo que observando los ejemplos lo entenderéis mejor.

Tabla 2. Ejemplo de conceptos operacionalizados en variables y preguntas de cuestionario

Concepto	Variable	Pregunta
Religiosidad	Asistencia a actos religiosos/ Número de veces que asiste a actos religiosos	¿Asiste a actos religiosos? / ¿En el último mes, a cuántos actos asistió aproximadamente?
Rendimiento académico	Nota media del curso anterior	¿Cuál fue su nota media en el curso anterior?
Ideología política	Recuerdo de voto/ Escala ideológica	¿A qué partido votó en las últimas elecciones? / Del 0 al 10 siendo 0 extrema izquierda y 10 extrema derecha…
Arraigo a la ciudad	Número de años que lleva viviendo en la ciudad	¿Cuántos años lleva viviendo en esta ciudad?

Viendo los ejemplos os surgirán preguntas como si realmente existe una sola forma de operacionalizar los conceptos. No, la realidad es que el investigador debe operacionalizar de la manera que mejor se adecue a sus objetivos e intereses, por ejemplo, la religiosidad podría operacionalizarse como práctica (como se ha expresado en la tabla) pero también como creencia (es decir, medir el tipo de fe que tiene el individuo: ateo, cristiano, etc.). Por último, me he dejado en el tintero la operacionalización de conceptos complejos, ya que superan los objetivos de este curso, para ello os remito al ya clásico texto de Cea D'Ancona (1996) sobre operacionalización de conceptos.

4) Diseño del cuestionario

Determinadas y operacionalizadas las variables, pasamos a diseñar la herramienta de extracción de datos más usada en investigación cuantitativa: el cuestionario. Un cuestionario no es otra cosa que un conjunto de preguntas que recoge, de manera ordenada, todas las variables que son relevantes para nuestra investigación, y cuyo fin es pasárselo (administrar es el término que se usa en la jerga) a los individuos que queremos estudiar. El objetivo del cuestionario es extraer información sobre las características (variables) clave que nos interesa estudiar, ni más ni menos. Aunque el objetivo de este curso no es diseñar cuestionarios, es conveniente repasar algunas cuestiones sobre cómo hacerlo.

Respecto a la formulación de preguntas, estas deben cumplir las siguientes reglas (Cruz-Beltrán, 2001):

1. Las preguntas deben ser corroborativas, es decir, solo buscan contrastar tal o cual característica, sentimiento u opinión del individuo. No son un examen.
2. Las preguntas no deben ser indiscretas ni condicionar la respuesta. Si preguntamos por cuestiones espinosas (racismo, relaciones sexuales…) conviene realizar preguntas de control para verificar que el individuo no trata de mentir o si se está guiando por principios de deseabilidad social (nadie te dirá que es racista, al menos que sea un ferviente creyente en la teoría racial nazi).
3. Deben formularse de manera directa, en segunda persona del singular.
4. Y lo más importante, deben adaptarse al nivel cultural de la población a encuestar. Es decir, debemos considerar el nivel cultural y el estatus social para adaptar el lenguaje a los códigos lingüísticos y el mundo de sentido de los actores (Cicourel, 2011).

Las categorías de repuesta en un cuestionario cuantitativo conviene que constituyan un sistema cerrado de respuesta potenciales que deben cumplir las siguientes características:

- Ser exhaustivas, es decir, que recoja toda la variabilidad social existente, todo el rango de variación de respuestas potenciales. En otras palabras, que todo el mundo pueda contestar a la pregunta.
- Ser excluyentes, es decir, que el individuo pueda elegir solo una categoría de respuesta y no más que una.

- Ser precisas, esto significa que deben intentar recoger la información en el mayor nivel de medida posible. Por ejemplo, la edad es mejor preguntarla como número que como intervalos, ya que así recogemos una información más detallada (la edad exacta y no el grupo de edad).

Por último, sobre el cuestionario hay también reglas para su elaboración, aunque lógicamente variarán según los intereses del investigador. Basta con que apliquéis las siguientes, que son con las que yo trabajo:

- El cuestionario debe empezar con una introducción atractiva, que explique muy brevemente el objetivo de investigación, quién es el que lo diseña, para qué, y especialmente, que asegure a los individuos que sus respuestas son totalmente anónimas y confidenciales. Yo soy partidario de que esta introducción sea breve, que se adapte al nivel cultural del encuestado (que sea siempre divulgativa) y que sea agradable, lúdica.
- Las preguntas de identificación, como sexo, edad, ingresos, ocupación, deben ir al final.
- Se debe comenzar con preguntas que busquen en la medida de lo posible enganchar al encuestado, por ejemplo preguntando su opinión o sentimientos sobre el tema que analizamos. Las preguntas más chulas al principio.
- Si el cuestionario es muy extenso, dividir en secciones. Si utilizáis cuestionarios online os recomiendo que cada pregunta sea una pantalla independiente y que además figure una barra de progreso para que el encuestado vea como avanza (el burro y la zanahoria o la lógica del videojuego).
- Las preguntas deben ser las justas y necesarias. El objetivo es extraer información de las variables que nos interesan, no hay que cansar a los encuestados con cuestionarios extremadamente largos sin necesidad.

5) Determinación de la muestra

Diseñado el cuestionario debemos plantearnos una pregunta fundamental ¿a quién se lo vamos a administrar? La respuesta parece simple, a toda la población que queramos estudiar, si por ejemplo mi objetivo es “Conocer qué explica el rendimiento académico de los estudiantes de primero de Sociología de la UGR”, podría ir y encuestar al 100% de la población, al fin y al cabo son solo tres grupos de, a lo sumo, unas 60 personas cada uno. Pero ¿qué pasa si mi población es demasiado grande? Por ejemplo, si quiero conocer la emancipación de los jóvenes españoles o la intención de voto de los españoles. En el caso en el que la población sea demasiado grande debemos recurrir a una muestra, es decir, a un subconjunto de la población que sea representativo de la misma en términos estadísticos. No es objetivo de este curso trabajar las técnicas de muestreo por lo que, si a alguien le interesa, os recomiendo que consultéis los textos de Rodríguez-Osuna (2001, 2005) que tenéis en la bibliografía. En este punto solo quiero destacar una cuestión que es relativamente polémica. Si bien el muestreo es necesario para extraer información cuantitativa que sea extrapolable a toda la población, ¿qué ocurre si por restricciones temporales o económicas no puedo diseñar una muestra

representativa? La respuesta ortodoxa es simple, no puedes hacer la investigación, porque tus datos no serían representativos de nada. Sin embargo, no estoy de acuerdo con la ortodoxia cuantitativa; si por restricciones impuestas no podemos acceder a construir una muestra representativa debemos igualmente iniciar la investigación y construir lo que se llama una muestra de conveniencia, es decir acceder a la población que se pueda, a veces incluso sin un criterio de muestreo previo. Los ortodoxos os dirán que esa muestra no representa nada de la población a estudiar, y en parte es cierto, sin embargo, sí que nos permite aproximarnos a la realidad con datos empíricos y, aunque estos no sean representativos de toda la población, sí que describen tendencias sociales que de otra manera no podríamos conocer. Eso sí, siempre debemos reconocer las limitaciones de manera honesta. Un ejemplo de una muestra de conveniencia lo podéis encontrar en el trabajo que realizamos Duque-Calvache, Torrado y Mesa-Pedrazas (2021) mediante una encuesta online para conocer la movilidad durante la pandemia (mirad los anexos online). Al fin y al cabo, ¿qué es mejor? ¿No hacer nada? ¿O aproximarnos como podamos a la realidad?

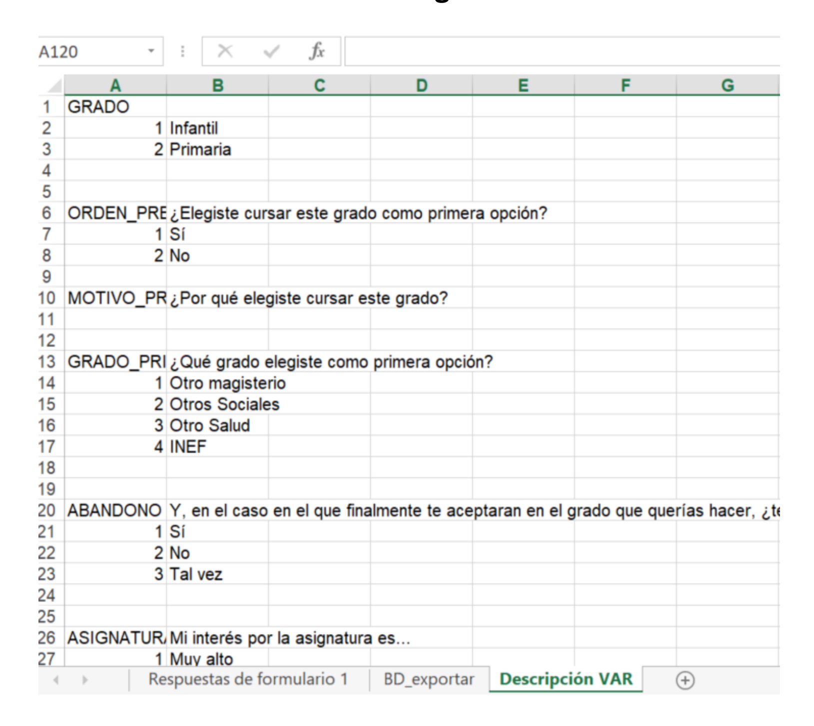
6) Trabajo de campo (Administración del cuestionario)

Diseñado el cuestionario y determinada la población que va a formar parte de la muestra debemos pasar a administrar el cuestionario a dicha población. Existen muchas técnicas de administración, las más populares a día de hoy son el cuestionario presencial asistido por un encuestador (en el que el encuestador entrevista una a una a cada persona), el cuestionario online (auto-administrado de manera autónoma por el encuestado) y el cuestionario auto-administrado presencial (en el que el encuestador va a un lugar, por ejemplo a una clase, y reparte el cuestionario a todos los encuestados). Si bien elegir una u otra técnica de administración puede parecer una tarea sencilla, no es tal. La elección de la estrategia de administración de la encuesta es tremendamente compleja y dependerá de tres factores: el nivel cultural de los encuestados, la naturaleza de la investigación y las técnicas de muestreo. Por ejemplo, no es lo mismo pasar el cuestionario a personas analfabetas que a personas que saben leer y escribir. Igualmente, tampoco tiene la misma efectividad una estrategia u otra si el cuestionario trata sobre rendimiento académico o sobre relaciones sexuales, o si el encuestado debe localizar puntos en un mapa (la mayoría de la gente no sabe leer un mapa). La elección de la estrategia de administración es compleja y debe ser sopesada de manera sesuda por el investigador. No obstante, no me extenderé sobre este punto, ya que no es central para el curso.

7) Tabulación de resultados

Una vez administrado el cuestionario a toda la muestra seleccionada debemos pasar a tabular los resultados, de manera que sean procesables por algún paquete estadístico (SPSS, R, STATA, etc.). Para ello, debemos construir dos archivos separados (yo suelo hacerlo utilizando el programa MS Excel). Por un lado, un “libro de códigos” o “estructura de los datos” en los que de manera clara se detallen: “nombre de la variable”, “descripción de la variable”, “etiqueta de cada valor/categoría de la variable” y “número asignado a cada valor/categoría”. Un ejemplo de este archivo lo tenéis abajo.

Figura 1. Ejemplo de estructura de datos de una encuesta elaborada por el profesor con una clase de alumnos del grado en Educación Primaria

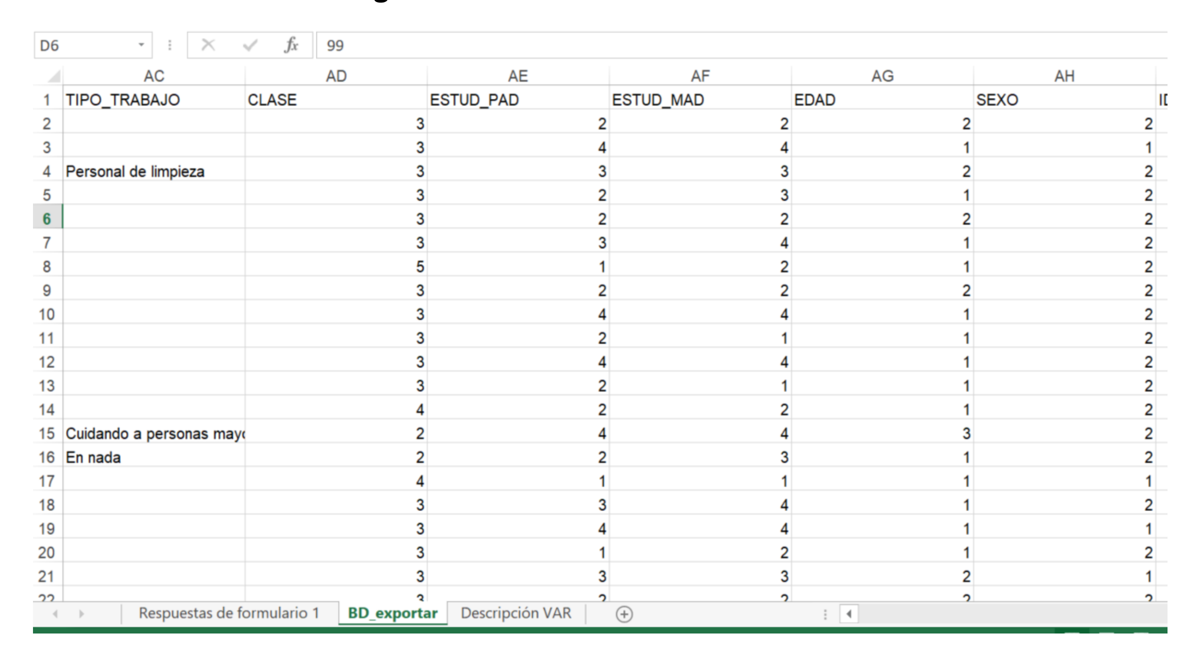


	A	B
1	GRADO	
2	1	Infantil
3	2	Primaria
4		
5		
6	ORDEN_PRE	¿Elegiste cursar este grado como primera opción?
7	1	Sí
8	2	No
9		
10	MOTIVO_PR	¿Por qué elegiste cursar este grado?
11		
12		
13	GRADO_PRI	¿Qué grado elegiste como primera opción?
14	1	Otro magisterio
15	2	Otros Sociales
16	3	Otro Salud
17	4	INEF
18		
19		
20	ABANDONO	Y, en el caso en el que finalmente te aceptaran en el grado que querías hacer, ¿te
21	1	Sí
22	2	No
23	3	Tal vez
24		
25		
26	ASIGNATUR	Mi interés por la asignatura es...
27	1	Muy alto

Respuestas de formulario 1 | BD_exportar | Descripción VAR

Adicionalmente, debemos construir una matriz de datos que solo contenga números, en la que cada fila sea un caso (individuo) y cada columna una variable (característica del individuo, pregunta del cuestionario). Los valores numéricos de la matriz deben corresponderse con el número asignado a cada valor/categoría que se encuentra en el archivo "estructura de datos". Un ejemplo de esto lo tenéis abajo.

Figura 2. Ejemplo de matriz de datos de una encuesta elaborada por el profesor con una clase de alumnos del grado en Educación Primaria



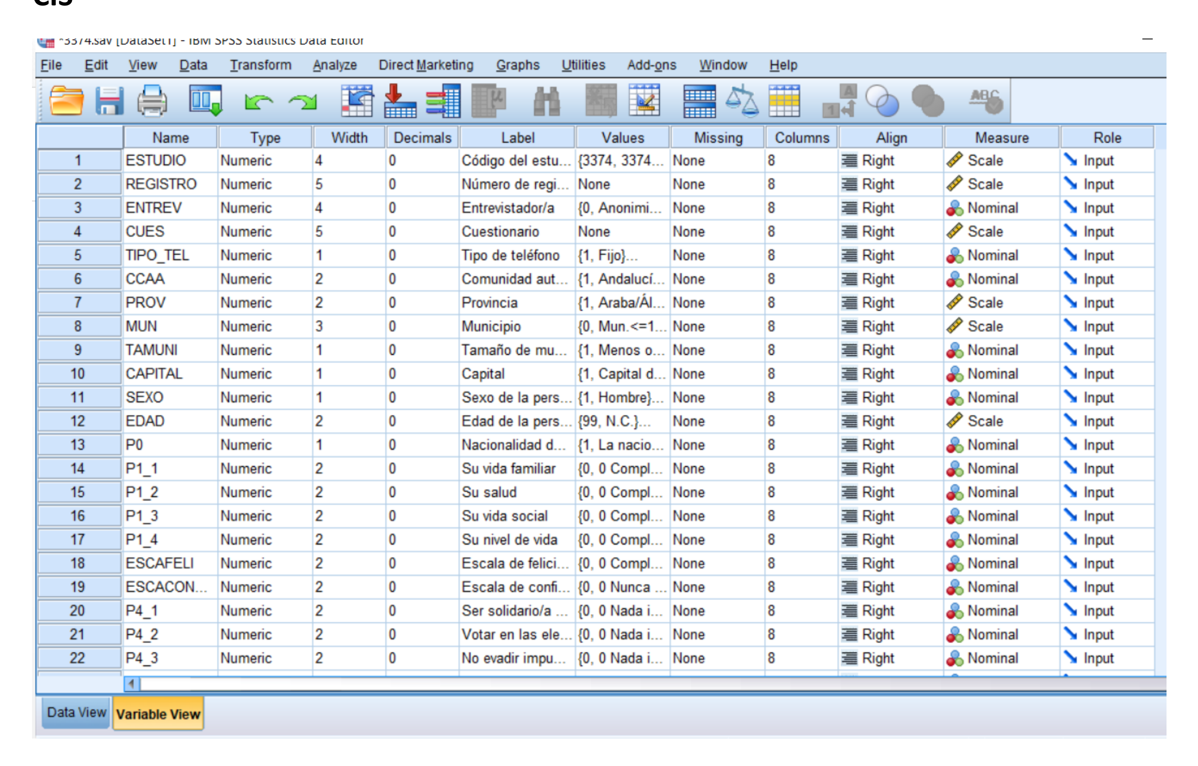
	AC	AD	AE	AF	AG	AH
1	TIPO_TRABAJO	CLASE	ESTUD_PAD	ESTUD_MAD	EDAD	SEXO
2		3	2	2	2	2
3		3	4	4	1	1
4	Personal de limpieza	3	3	3	2	2
5		3	2	3	1	2
6		3	2	2	2	2
7		3	3	4	1	2
8		5	1	2	1	2
9		3	2	2	2	2
10		3	4	4	1	2
11		3	2	1	1	2
12		3	4	4	1	2
13		3	2	1	1	2
14		4	2	2	1	2
15	Cuidando a personas may	2	4	4	3	2
16	En nada	2	2	3	1	2
17		4	1	1	1	1
18		3	3	4	1	2
19		3	4	4	1	1
20		3	1	2	1	2
21		3	3	3	2	1

Respuestas de formulario 1 | BD_exportar | Descripción VAR

Finalmente la matriz de datos se abre en un paquete estadístico, y a continuación se le añade la estructura de datos (el proceso para hacer esto es relativamente complejo y

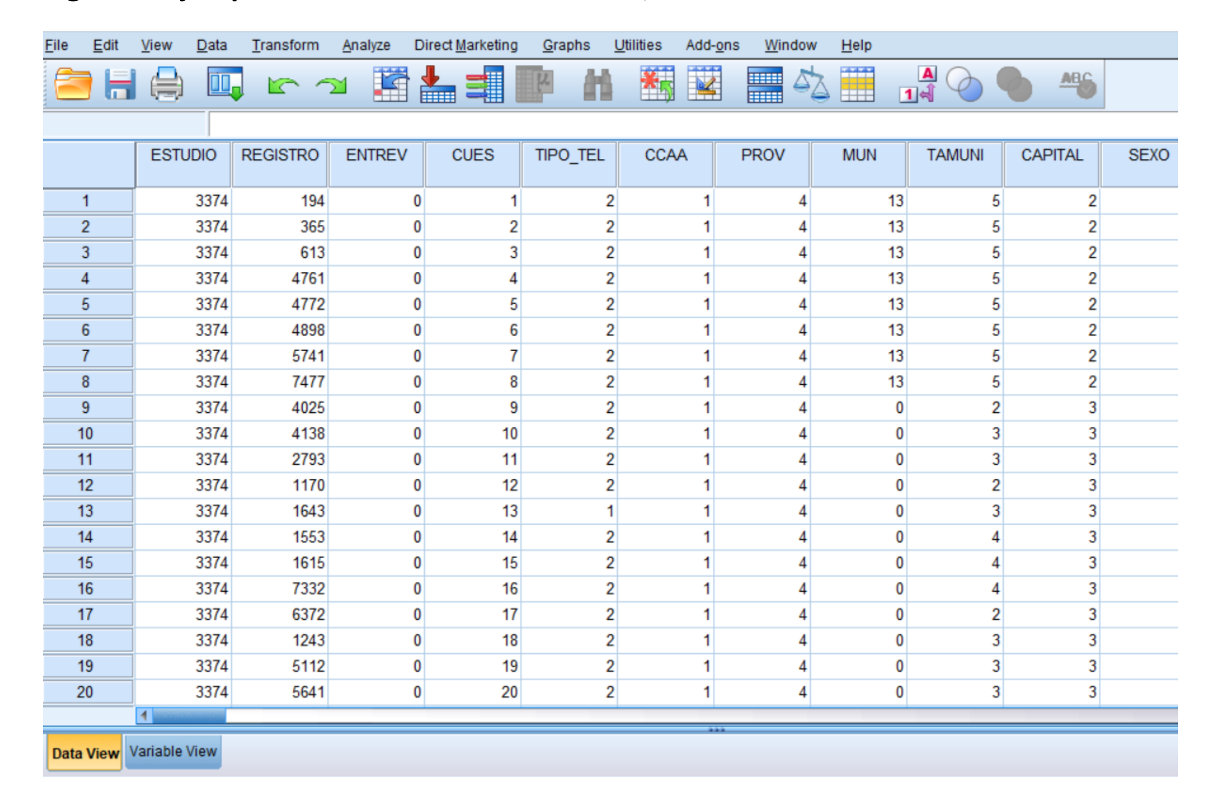
supera los objetivos de este curso, para más información contactar con el profesor). Por ejemplo, SPSS os permite ver la estructura de datos clicando en la pestaña "vista de variables" (figura 3) y la matriz de datos si clicáis en la pestaña "vista de datos" (figura 4).

Figura 3. Ejemplo de "vista de variables" de SPSS, de la "Encuesta Flash Senado" del CIS



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	ESTUDIO	Numeric	4	0	Código del estu...	{3374, 3374...	None	8	Right	Scale	Input
2	REGISTRO	Numeric	5	0	Número de regi...	None	None	8	Right	Scale	Input
3	ENTREV	Numeric	4	0	Entrevistador/a	{0, Anonimi...	None	8	Right	Nominal	Input
4	CUES	Numeric	5	0	Cuestionario	None	None	8	Right	Scale	Input
5	TIPO_TEL	Numeric	1	0	Tipo de teléfono	{1, Fijo}...	None	8	Right	Nominal	Input
6	CCAA	Numeric	2	0	Comunidad aut...	{1, Andalucí...	None	8	Right	Nominal	Input
7	PROV	Numeric	2	0	Provincia	{1, Araba/Ál...	None	8	Right	Scale	Input
8	MUN	Numeric	3	0	Municipio	{0, Mun.<=1...	None	8	Right	Scale	Input
9	TAMUNI	Numeric	1	0	Tamaño de mu...	{1, Menos o...	None	8	Right	Nominal	Input
10	CAPITAL	Numeric	1	0	Capital	{1, Capital d...	None	8	Right	Nominal	Input
11	SEXO	Numeric	1	0	Sexo de la pers...	{1, Hombre}...	None	8	Right	Nominal	Input
12	EDAD	Numeric	2	0	Edad de la pers...	{99, N.C.}...	None	8	Right	Scale	Input
13	P0	Numeric	1	0	Nacionalidad d...	{1, La nacio...	None	8	Right	Nominal	Input
14	P1_1	Numeric	2	0	Su vida familiar	{0, 0 Compl...	None	8	Right	Nominal	Input
15	P1_2	Numeric	2	0	Su salud	{0, 0 Compl...	None	8	Right	Nominal	Input
16	P1_3	Numeric	2	0	Su vida social	{0, 0 Compl...	None	8	Right	Nominal	Input
17	P1_4	Numeric	2	0	Su nivel de vida	{0, 0 Compl...	None	8	Right	Nominal	Input
18	ESCAFELI	Numeric	2	0	Escala de felici...	{0, 0 Compl...	None	8	Right	Nominal	Input
19	ESCACON...	Numeric	2	0	Escala de confi...	{0, 0 Nunca ...	None	8	Right	Nominal	Input
20	P4_1	Numeric	2	0	Ser solidario/a ...	{0, 0 Nada i...	None	8	Right	Nominal	Input
21	P4_2	Numeric	2	0	Votar en las ele...	{0, 0 Nada i...	None	8	Right	Nominal	Input
22	P4_3	Numeric	2	0	No evadir impu...	{0, 0 Nada i...	None	8	Right	Nominal	Input

Figura 4. Ejemplo de "vista de datos" de SPSS, de la "Encuesta Flash Senado" del CIS



	ESTUDIO	REGISTRO	ENTREV	CUES	TIPO_TEL	CCAA	PROV	MUN	TAMUNI	CAPITAL	SEXO
1	3374	194	0	1	2	1	4	13	5	2	
2	3374	365	0	2	2	1	4	13	5	2	
3	3374	613	0	3	2	1	4	13	5	2	
4	3374	4761	0	4	2	1	4	13	5	2	
5	3374	4772	0	5	2	1	4	13	5	2	
6	3374	4898	0	6	2	1	4	13	5	2	
7	3374	5741	0	7	2	1	4	13	5	2	
8	3374	7477	0	8	2	1	4	13	5	2	
9	3374	4025	0	9	2	1	4	0	2	3	
10	3374	4138	0	10	2	1	4	0	3	3	
11	3374	2793	0	11	2	1	4	0	3	3	
12	3374	1170	0	12	2	1	4	0	2	3	
13	3374	1643	0	13	1	1	4	0	3	3	
14	3374	1553	0	14	2	1	4	0	4	3	
15	3374	1615	0	15	2	1	4	0	4	3	
16	3374	7332	0	16	2	1	4	0	4	3	
17	3374	6372	0	17	2	1	4	0	2	3	
18	3374	1243	0	18	2	1	4	0	3	3	
19	3374	5112	0	19	2	1	4	0	3	3	
20	3374	5641	0	20	2	1	4	0	3	3	





8) Análisis de datos

Teniendo la matriz de datos lista podemos pasar a la parte que más nos interesa en este curso, el análisis de datos. Aunque depende de los autores que consultéis y de las preferencias de cada investigador, de cara a este curso vamos a considerar que existen tres fases generales en el análisis de datos: el análisis descriptivo univariable, el análisis bivariable y el análisis inferencial (el análisis multivariante lo excluiremos de este curso introductorio).

El **análisis descriptivo univariable**, como su nombre indica tiene por objetivo obtener una descripción global de la base de datos que hemos construido. Para ello, se recurre a una técnica estadística clásica, el análisis de tablas de frecuencias. Las tablas de frecuencias son tablas que resumen la información de una sola variable, indicando cuántos casos se aglutinan en cada una de sus categorías o valores, dicho de otro modo, resumen cuántos individuos han respondido a cada una de las categorías de respuesta posibles en una pregunta del cuestionario. En la tabla 3 podéis ver la distribución de frecuencias de la variable “recuerdo de voto en las últimas elecciones” de la encuesta “Flash Debate Senado” realizada por el Centro de Investigaciones Sociológicas (CIS) en octubre de 2022.

Tabla 3. Tabla de frecuencias de la variable “recuerdo de voto en las anteriores elecciones”

	f
PP	160
PSOE	217
Ciudadanos	48
En Comú Podem	17
Més Compromís	4
ERC	22
JxCat	11
EAJ-PNV	15
EH Bildu	8
CC-PNC-NC	1
Navarra Suma (UPN)	4
PACMA	1
VOX	52
CUP	4
Unidas Podemos	65
BNG	2
PRC	1
Más País	7
En Común-Unidas Podemos	3
Teruel Existe	1
Voto nulo	4
Otros partidos	9
En blanco	17
No recuerda	51
N.C.	116
Total	837
N.P.	175
Total	1012

Fuente: CIS 2022, Encuesta Flash Senado

El análisis de tablas de frecuencias es fundamental como paso previo en cualquier investigación cuantitativa, en tanto que nos permite tener una descripción global de las variables que contiene nuestra base de datos, concretamente de cómo son (¿Qué ha respondido la gente?), la cantidad de datos que se pierden (¿Hay muchas no respuestas?) y de reflexionar sobre una cuestión fundamental, la recodificación.

En ocasiones las tablas de frecuencias pueden ser muy grandes y complejas por contener demasiada información que a lo mejor no es necesaria para nuestra investigación, por ello es necesario recodificarlas. Recodificar no es otra cosa que agrupar las categorías en un número menor para poder simplificar la información y que sea más legible. Por ejemplo, la tabla 3 resulta poco intuitiva y su lectura es difícil, por ello podríamos optar por recodificarla en menos categorías, por ejemplo, en las que vemos en la tabla 4.

Tabla 4. Tabla de frecuencias de la variable "recuerdo de voto en las anteriores elecciones", recodificada por el profesor

	f
PP	159
PSOE	217
C's	48
Podemos	85
VOX	52
Nacionalistas/Regionalistas	48
Otros partidos	41
Abstención, nulo, blanco	180
N.R./N.C	182
Total	1012

Fuente: CIS 2022, Encuesta Flash Senado

Ya veremos más adelante cómo obtener tablas de frecuencias y cómo recodificar en SPSS, por ahora baste una advertencia. Para recodificar es necesario atender a un doble criterio: primero, es necesario que las nuevas categorías (que agrupan a las antiguas) tengan un sentido teórico, por ejemplo, en la tabla 4 no hubiera tenido ningún sentido agrupar al PP con ERC (Esquerra Republicana de Catalunya). Segundo, hay que respetar un criterio estadístico, usualmente es preferible agrupar categorías pequeñas, con pocos casos, y dejar las que tienen más casos aparte. No obstante, esto siempre dependerá del criterio experto del investigador.

Este ejercicio de analizar tablas de frecuencias y recodificar cuando se considere necesario es el paso primero y fundamental del proceso de análisis de datos cuantitativos, y debe realizarse con todas las variables (una a una) que queramos a analizar.

Por su parte, una vez hemos obtenido una visión global de las variables a analizar y hemos recodificado aquellas que sean necesarias, debemos pasar a la segunda fase: **el análisis bivariable**. El análisis bivariable consiste en cruzar las variables, para ver cómo se distribuye una opinión o comportamiento entre los distintos grupos sociales. Siempre debéis introducir vuestra variable a explicar o dependiente en las columnas, y las explicativa en las filas. Por ejemplo, en la tabla 5, tenéis el cruce de las variables

"frecuencia de asistencia a actos religiosos" (variable dependiente o a explicar) y "sexo" (variable independiente o explicativa), en ella lo que podéis apreciar es cómo se distribuyen los casos en cada una de las casillas (por ejemplo hay 180 hombres que nunca asisten a actos religiosos).

Tabla 5. Tabla de contingencia de las variables "frecuencia de asistencia a actos religiosos" y "sexo". Números absolutos

	Nunca	Casi nunca	Varias veces al año	Dos o tres veces al mes	Todos los domingos y festivos	Varias veces a la semana	N.C.	Total
Hombre	180	159	121	42	69	23	7	601
Mujer	208	179	191	63	116	42	9	808
Total	388	338	312	105	185	65	16	1409

Fuente: CIS 2022, Encuesta Flash Senado

Ahora bien ¿cómo podemos interpretar esta tabla? Podemos hacer dos cosas (en realidad tres, pero vamos a dejarlas en dos). Uno, analizar los porcentajes sobre la fila, lo que podría interpretarse como la propensión o probabilidad que tienen hombres y mujeres de asistir a actos religiosos. Por ejemplo, viendo la tabla 6 podríamos decir que la probabilidad que tienen los hombres de no asistir nunca o casi nunca a actos religiosos es del 56% (30% nunca + 26% casi nunca), mientras que en las mujeres es del 48% (26% nunca + 22% casi nunca), por lo que podríamos deducir que las mujeres son más practicantes que los hombres, o viceversa, que los hombres son menos practicantes que las mujeres.

Tabla 6. Tabla de contingencia de las variables "frecuencia de asistencia a actos religiosos" y "sexo". Porcentaje de filas

	Nunca	Casi nunca	Varias veces al año	Dos o tres veces al mes	Todos los domingos y festivos	Varias veces a la semana	N.C.	Total
Hombre	30%	26%	20%	7%	11%	4%	1%	100%
Mujer	26%	22%	24%	8%	14%	5%	1%	100%
Total	28%	24%	22%	7%	13%	5%	1%	100%

Fuente: CIS 2022, Encuesta Flash Senado

O, dos, podríamos analizar los porcentajes sobre el total de cada columna, que pueden interpretarse como el perfil de las frecuencias de prácticas religiosas. Si observamos la tabla 7 vemos como el perfil de los que nunca practican su religión es 54% de mujeres y 46% de hombres. Esto podría daros a pensar que ocurre justo lo contrario que habíamos dicho, que las mujeres son menos practicantes, pero esto es un error de interpretación. Lo que indica es simplemente que hay más mujeres en la categoría, nada más. Y que haya más mujeres se debe a que, si os fijáis, hay globalmente más en el total de la muestra (57% del total de la muestra son mujeres). Por tanto, recordad: si queréis analizar perfiles usad porcentaje de columnas, si queréis analizar propensiones, de filas. Introduciendo siempre la variable a explicar (en este caso "frecuencia de asistencia actos religiosos") en las columnas y la explicativa o independiente (en este caso "sexo") en las filas.

Tabla 7. Tabla de contingencia de las variables "frecuencia de asistencia a actos religiosos" y "sexo". Porcentaje de columnas

	Nunca	Casi nunca	Varias veces al año	Dos o tres veces al mes	Todos los domingos y festivos	Varias veces a la semana	N.C.	Total
Hombre	46%	47%	39%	40%	37%	35%	44%	43%
Mujer	54%	53%	61%	60%	63%	65%	56%	57%
Total	100%	100%	100%	100%	100%	100%	100%	100%

Fuente: CIS 2022, Encuesta Flash Senado

Por último, el último paso en el análisis de datos consiste en **inferir**. Pero ¿qué es eso de inferir? Inferir consiste en extrapolar las conclusiones extraídas de una muestra a la población en su conjunto. Si os parece difícil recordad que vosotros hacéis inferencias constantemente, por ejemplo, cuando vais a tomar unas cervezas, si veis un pub en el que la gente de la puerta lleva chaqueta, inferís que el pub es para gente "de etiqueta" o "pija". Pues esto es lo mismo, consiste en extraer conclusiones de nuestra muestra, y extrapolarlas al conjunto de la población. ¿Cómo se hace esto? En estadística utilizamos lo que se llaman pruebas de contraste de hipótesis, siendo la más común la prueba del Chi-cuadrado. Esta prueba nos dice si la correlación entre dos variables es extrapolable a la población (es significativa) o no lo es (no es significativa). En SPSS se puede solicitar esta prueba siempre que crucemos dos variables (luego veremos cómo) y nos da la siguiente tablita:

Tabla 8. Prueba de contraste Chi-Cuadrado en SPSS de la tabla de contingencia entre las variables "frecuencia de asistencia a actos religiosos" y "sexo"

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	10,673[a]	6	,099
Likelihood Ratio	10,712	6	,098
Linear-by-Linear Association	6,788	1	,009
N of Valid Cases	1409		

Fuente: CIS 2022, Encuesta Flash Senado

¿Cómo se interpreta esto? Es sencillo. Si la significación asintótica (Asymptotic Significance) del "Pearson Chi-Square" es menor o igual que 0,05 entonces diremos que la correlación entre las variables es significativa, y en nuestro caso, que el sexo explica la frecuencia de asistencia a actos religiosos. Si ese valor es mayor que 0,05 entonces diremos que no es significativa, y por tanto no podremos extrapolar las interpretaciones que hagamos de la tabla (como lo que decíamos de que las mujeres son más practicantes) al conjunto de la población, como es nuestro caso, al ser ese valor de 0,099, o sea, mayor que 0,05 y por tanto, no significativo.





9) Selección y síntesis de los resultados más relevantes

Una vez realizado el análisis, el investigador debe seleccionar aquellos resultados obtenidos que son centrales para contestar a su objetivo de investigación. Para ello, además, debe sintetizarlos en unas pocas tablas o gráficos que sean de fácil lectura para el público potencial. Para ello suelen utilizarse varias opciones; o bien optar por tablas grandes que sinteticen información, o bien construir representaciones gráficas de diversa índole, como histogramas, gráficos de barras, etc.

10) Redacción de artículo

Por último, completado el proceso, la última fase es reflejarlo en un escrito que recoja los principales resultados. Aunque existen muchos formatos en los que presentar nuestra investigación, el más extendido en la actualidad es el "artículo científico". Un artículo científico tiene las siguientes partes:

- Introducción. En la que el autor muestra al público la importancia del artículo, justificando la pertinencia de su objetivo de investigación, explicitando dicho objetivo y explicando de manera general una síntesis del artículo.
- Marco teórico. En este apartado el autor debe sintetizar en un texto propio las principales aportaciones previas que han realizado otros autores que han tratado su problema de investigación. Es fundamental este marco para extraer hipótesis y para finalmente enfrentarlo a los resultados que nosotros obtenemos.
- Metodología. Debe contener toda la información metodológica fundamental referente a las técnicas de extracción de datos (fuente o cuestionario diseñado, población y muestra, variables clave) y técnicas de análisis de datos empleada.
- Resultados. Aquí expondremos de manera ordenada y clara la información más relevante seleccionada y procesada en el paso 9) que detallamos en este documento.
- Conclusiones. Sintetizamos nuestros principales hallazgos, sus implicaciones y los enfrentamos a la literatura previa revisada en el marco teórico.
- Bibliografía. Incluiremos un apartado, no numerado, en el que consten las citas completas a todas las obras citadas en nuestro texto.

Bibliografía

Cea D'Ancona, M.A. (1996). La Operacionalización de conceptos. En Cea D'Ancona, M.A., *Metodología cuantitativa. Estrategias y Técnicas de Investigación Social*, Madrid: Síntesis.

Cruz-Beltrán, F. (2001). El diseño de cuestionario. En Gualda Caballero, E. y Cruz Beltrán, F., *Métodos y técnicas de investigación social*. Huelva: Grupo de Investigación Estudios Sociales e Intervención Social.

Cicourel, A. (2011). Método y medida en Sociología. Madrid: Centro de Investigaciones Sociológicas.



Duque-Calvache, R., Torrado, J. M., y Mesa-Pedrazas, Á. (2021). Lockdown and adaptation: residential mobility in Spain during the COVID-19 crisis. European Societies, 23(sup1), S759-S776. https://doi.org/10.1080/14616696.2020.1836386

Macionis, J. y Plummer, K. (2007). *Sociología*. Madrid: Pretince Hall

Rodríguez Osuna, J. (2001). *Métodos de muestreo*. Madrid: Centro de Investigaciones Sociológicas.

Rodríguez Osuna, J. (2005). *Métodos de muestreo. Casos prácticos*. Madrid: Centro de Investigaciones Sociológicas.